C.P.

Pvp

A ROTATION SAMPLING PLAN FOR ESCS QUARTERLY HOG SURVEYS

C. P. Gleason
Statistical Research Division
Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture
Washington, D. C. 20250

April 1979

A ROTATION SAMPLING PLAN FOR ESCS QUARTERLY HOG SURVEYS. By C. Gleason; Statistical Research Division; Economics, Statistics, and Cooperatives Service; U.S. Department of Agriculture; Washington, D.C. 20250; April 1979.

ABSTRACT

A rotation sampling plan to estimate hog inventories is presented and a composite double sampling regression or ratio estimator is compared to a single time direct expansion estimator.

Key words: Rotation sampling, successive sampling, panel studies, regression or ratio estimation, change estimation, double sampling.

CONTENTS

SUMMARY

An analysis of different rotation designs to be used with the list frame in the Multiple Frame Quarterly Hog Survey is presented and analyzed. The rotation design discussed deals only with a four-quarter survey period and does not address between year rotation plans. A combined ratio estimator, separate ratio, and separate regression estimator are compared to the direct expansion estimator.

A rotation design where 50 percent of the sample units are matched between quarters is suggested from the analysis. The rotation plan suggested is:

Dec	March	June	Sept
S_1			S_1
S_2	S_2		
	S_3	S_3	
		S_4	S_4

where S_i is a replication of size 1/2 the stratum allocation.

This design allows for estimation of current inventory with the capability of more reliable estimates of change to be developed. The design also reduces respondent burden to two contacts per year in nearly all strata.

INTRODUCTION

Current procedures: Currently, ESCS uses several different rotation sampling plans for the list frame in the Hog Multiple Frame Survey. All rotation plans discussed in this paper deal with sampling within a year for quarterly estimates in December, March, June and September. This list frame is updated annually before selection of the sample(s) to be used for the quarterly surveys. The problem of rotating the sample between years given an annual update is reserved for a later paper.

Figure 1 gives the types of rotation plans currently used within strata by states on the quarterly hog survey. In the strata below extreme operators, several states select one sample from their list for use during the entire survey period of four quarterly surveys. This gives 100 percent overlap between quarters as illustrated by Plan 1 in Figure 1. In other states a new sample is selected midway within the year and there is no overlap in the sample between quarters (Plan 2, Figure 1). Note that Plan 3 is a simple variation of Plan 2, but an individual operator would be contacted at six-month intervals. In some E.O. strata, ESCS has a policy of no overlap between survey periods. This rotation design (Plan 4) minimizes the respondent burden for larger operations to one contact for an entire year. Each state may use a combination of each of the above plans.

Figure 1: Current Rotation Plans used on ESCS Hog Quarterly Survey

Plan 1: One Sample, held in 4 times, 100% overlap (OL) between periods

Dec	March	June	Sept
S_1	S_1	S_1	S_1

Plan 2: Two samples, each held in two periods, 100% OL between Dec/March, 0% OL between March/June

Dec	March	June	Sept
S_1	S_1		
		S_2	S_2

Plan 3: Two samples, each held in two periods, 100% OL between Dec/June 0% OL between Dec/March, March/June, June/Sept

Dec	March	June	Sept
S_1		S_1	
	S_2		S_2

Plan 4: Four samples, 0% OL between survey periods

Dec	March	June	Sept
S_1			
	S_2		
		S_3	
			S_4

Table A1 in Appendix A gives the 1978 rotation plan for the 14 hog multiple frame states. It can be seen in Table A1 that a multitude of different rotation plans are being used in the quarterly hog survey. No consistent overall policy on rotation for the quarterly hog surveys exists at present.

Estimation of Change: The use of successive sample surveys of the same population provides flexibility in choosing a sampling design. Different estimators can be utilized to supplement a direct expansion estimate based only on current survey data. Repetitive sampling

provides the necessary data to form estimates based on change. Estimates of survey to survey changes are made by matching reports from the same sampling unit surveyed in two or more successive surveys.

Using matched reports as a basis for estimation has a long history with ESCS -- Statistics. This approach continues to be heavily relied upon for nonprobability surveys. An indication called the "current/current" ratio is calculated and applied to previous survey estimates to provide a current survey estimate. A variation of the "current/current" indication is the "current/historic" ratio. The latter indication also measures change from a previous survey to the current period, but the data for the prior period are collected on the current questionnaire. The advantage is that no prior survey needs to be conducted, therefore, reducing respondent burden. This would be particularly useful when a statistical series is being initiated since prior data for sample units in the current survey would not be available. For an ongoing program, however, it has been found that data reported for the preceding period are often subject to memory bias [4].

For probability surveys, estimates can be made without prior survey data for matched sample units. Direct expansion estimates have been given primary emphasis for our June Enumerative and Multiple Frame Livestock survey. This does not mean that one should ignore available prior information.

Ratios and ratio estimators have been computed based on matched segments in the June and December Enumerative surveys. Extreme

operators have been excluded from the matching procedure. The estimated sampling error associated with the ratio estimators has been about the same as those associated with the direct expansion estimate. Therefore, the ratio estimator has not been emphasized. No attempt has been made to form a composite estimate based on combining the ratio estimate from matched segments with a direct expansion for the unmatched "rotated in" segments.

Little use has been made of matching reports from multiple frame surveys, even though the same sample unit is often surveyed in two or more successive surveys. Cochran [4, p. 342] states that if we wish to maximize precision, the following can be said about replacement policy:

"1. For estimating change, it is best to retain the same sample throughout all occasions.

2. For estimating the average over all occasions, it is best to draw a new sample on each occasion.

3. For current estimates, equal precision is obtained either by keeping the same sample or by changing it on every occasion. Replacement of part of the sample on each occasion may be better than these alternatives."

Since we are not interested in estimating the average number of hogs in a calendar year, statement two above does not apply to our situation. We will show that a partial rotation plan of approximately 50% matched units between occasions is well suited for estimation of hog inventories.

THE ROTATION PLAN

Use of successive sampling or rotation sampling in surveys has been investigated by many authors for several types of surveys [1,2,5,6,8,10,11,12,14]. Gleason and Tortora[8] have extended the successive sampling theory for use with two overlapping sampling frames, as in ESCS livestock multiple frame surveys. In the analysis which follows three different successive sampling estimators will be compared in terms of precision to the direct expansion estimate on a stratum by stratum basis and then for estimation of the population total for the list. The estimators are:

1. The ratio estimate within strata.
2. The regression estimate within strata.
3. The combined ratio estimate over all strata.

Appendix B gives the appropriate mathematical and statistical expressions for the above estimators and their variances.

The list portion of the multiple frame data file was reformatted, and merged together by stratum and operator ID for each data set. An approximate 50% sample within strata was selected from all operators who responded in both survey periods. The sample was selected to simulate a rotation design of approximately 50% matching of units between survey periods. This matching will allow us to develop ratio and regression estimators and estimate optimum matching percentage between survey periods. Table A3 in Appendix A gives the number of matched samples selected by stratum and state.

Figure 2 below depicts the rotation design which will be investigated.

Figure 2

Dec	March	June	Sept
S_1			S_1
S_2	S_2		
	S_3	S_3	
		S_4	S_4

S_i is a random sample of size ½ of those who responded both times.

Note that for this rotation design only two contacts of each operation per year is made and this design permits composite estimation using matched and unmatched reports.

In the analysis sections which follow we will investigate the list frame strata in the following states and survey periods:

Iowa 1978 December, March

Iowa 1976 June, September

Minnesota 1977 June, September

Kansas 1977 December, March

PORTION OF SAMPLE TO BE RETAINED

One of the major considerations in an ongoing survey is what percentage of the sample should be retained from one survey period to the next. Since the quarterly hog survey is a stratified design the optimum percentage to match varies by stratum. Table 1 gives the optimum percent to match for each data set analyzed. The variable considered was "total hogs and pigs." The appropriate mathematical expression for estimation of the optimum matching

Table 1: Correlation coefficients between survey periods for total hogs and optimum percentage to match, by state, stratum.

Iowa Dec/March 1978

Stratum	Correl Coeff	Optimum % to Match reg est	ratio est
2	.38	.49	.54
3	.36	.51	.53
4	.77	.41	.41
5	.81	.41	.42
6	.67	.44	.44
806	.93	.27	.29

Iowa June/Sept 1976

Stratum	Correl Coeff	Optimum % to Match reg est	ratio est
3	.70	.42	.43
4	.78	.39	.41
5	.74	.43	.44
6	.84	.37	.38
7	.86	.36	.36

Minnesota June/Sept 1977

Stratum	Correl Coeff	Optimum % of Match reg est	ratio est
11	.99	.06	.14
12	.96	.22	.22
31	.28	.47	.54
61	.83	.34	.37
62	.79	.37	.38
63	.85	.36	.36
64	.74	.41	.42
65	.82	.40	.41

Kansas Dec/Mar 1977

Stratum	Correl Coeff	Optimum % of Match reg est	ratio est
10	.85	.34	.35
12	.80	.37	.39
14	.75	.40	.40
15	.74	.41	.42
16	.73	.42	.44

percentage for both a regression or ratio estimator is provided in Appendix B. The optimum percentage for the different types of estimators is a function of the correlation coefficient between survey periods for the same variable.

The optimum percentage to match is very similar for the regression and the ratio estimates. The optimum percentage to match for both estimators is generally between 30 and 50 percent. Fourty-one (41) percent is the median of the percentages to optimally match over all states and strata analyzed. Therefore, assuming an overall 80 percent response rate, 50 percent of the units should be retained between survey periods by stratum to operationally have a 40 percent matched sample between survey periods.

Also, inspecting Table 1 for Iowa we see that stratum correlation coefficients do not remain constant between years or different survey periods. It is planned to investigate this further for an entire four quarter survey period.

In the next section we will compare the various estimators with respect to their variances.

COMPARISON OF ESTIMATORS

The three composite estimators which are to be compared to the direct expansion estimator are presented in Appendix B. Basically the composite estimator is of the form:

$$\hat{Y} = W\,\hat{Y}_m + (1 - W)\,\hat{Y}_u \qquad (1)$$

where $\hat{Y}_m$ is a ratio or regression estimate of the total from the match portion of the sample, $\hat{Y}_u$ is an estimate of the total from the unmatched

portion of the sample and W is a weight used to combine the two independent estimates. $\hat{Y}_m$ can be a combined estimate over all strata which have matching units or might consist of separate ratio or regression estimators for each stratum. The mean/unit, regression and ratio estimates for each state and stratum appear in Table A4 in Appendix A.

Table 2 gives the relationship between estimates of the total and the coefficient of variation for each estimator. Note that the direct expansion estimate is considered as the base estimate with each of the other estimators written as a percentage of the direct expansion. In most cases the composite estimators are within 2 percent of the direct expansion except for the combined ratio estimate in Minnesota.

Table 2: Estimate of total and for direct expansion, regression, separate ratio and combined ratio estimators

State	D.E. Base % (%)	D.E. C.V. (%)	Reg % of D.E. (000)	Reg C.V. (%)	Separate Ratio % of D.E. (000)	C.V. (%)	Combined % of D.E. (000)	Combined Ratio C.V.
Iowa March (1978)	100.0	3.3	100.0	3.2	101.0	3.2	100.0	3.3
Iowa Sept (1976)	100.0	3.6	101.1	3.3	101.5	3.3	*	*
Minnesota Sept (1977)	100.0	4.2	101.3	3.8	101.4	3.8	101.8	3.8
Kansas March (1977)	100.0	11.4	100.9	11.3	101.2	11.2	101.3	11.2

* not available

Table 2 also shows that the precision of the estimate was slightly improved when using a regression or ratio estimate over a direct expansion estimate. In addition, the successive sampling estimators may be more stable since a portion of the sample is retained from quarter to quarter and the estimates are linked through the use of the composite estimator.

Table 3 shows in percentage terms the gain in precision on the variance of the respective estimators over the direct expansion. Gains are modest in most cases. Little gain was realized for Kansas since the composite estimators were used only in strata 14, 15, and 16, and direct expansion estimates in the other strata.

Table 3: Gain in precision (variance of the estimate) for regression, ratio, and combined ratio estimators over the direct expansion estimator

State	Gain Reg Over D.E.	Gain Separate Ratio Over D.E.	Gain Combined Ratio Over D.E.
	%	%	%
Iowa March (1978)	10.1	6.2	.5
Iowa Sept (1976)	12.7	12.0	*
Minnesota	22.9	17.4	16.0
Kansas	.3	.3	.3

*not available

Note that in this development we were interested only in current estimates of level. Efficient estimators of quarter to quarter change should also be investigated. Other government agencies are already using rotation designs to estimate current level and change for their operational data series [3,9,15,16]. However, no optimal solution exists to estimate both change and level simultaneously. Further theoretical research needs to be done to attack this

important survey problem. This problem can be simply stated as follows: The optimum estimator of change added to the previous quarter's estimate does not necessarily equal the best estimate of the current quarter's level.

One of the most important aspects of using a composite estimator is choosing the weight (W) to combine the matched and unmatched portion of the sample for the composite estimator. Good sampling design practice dictates that this quantity should be fixed prior to estimation and not estimated from the sample used [4]. However, when several characteritics are being estimated such as total hogs, farrowings and breeding stock, etc., these weights are going to vary for each characteristic and list frame stratum. It would be extremely difficult to estimate an optimum weight for each characteristic and stratum. Therefore, it is desirable to fix W prior to conducting the survey. The range of optimum weights for each type of estimator as computed from the surveys studied may be seen in Table A6 in Appendix A; in most cases the weight is nearly one-half. Based on these results and the fact that any weight W provides an unbiased estimator, it is recommended that a weight of .50 be used to combine the unmatched and matched portion of the sample in a composite regression or ratio estimator.

A question also arises for the regression estimator about the values of the regression coefficients β_o and β_1 and their estimation from the matched sample. The β_o coefficient varies considerably by strata and no definite fixed value can be suggested from the analysis. The β_1 term in the estimated regression does not vary as much as the β_o term, but fixing a value for estimation over all strata for all characterstics at this point is a bit premature. However, Cochran [4, p. 191] shows that such a fixed β_1 value

leads to an unbiased estimate of the mean. Since the estimates of β_o and β_1 varied considerably by strata it was felt that a combined regression estimate would not be appropriate in this situtation [4, p. 203 and 204] . The dangers of bias with small sample sizes (as in Minnesota Strata 11) may also put the separate regression estimate into an unfavorable category.

It seems clear that if one is estimating the paramters of the model

$$y = \beta_o + \beta_1 x,$$

where y is the number of hogs for the current quarter and x is the previous quarter's total hogs that β_o should be o and β_1 would be just a percentage decrease or increase from the previous month's level, i.e., a ratio estimate would be appropriate. Inspecting Table A7 in Appendix A we see that for most state/stratum combinations that we do not reject the null hypothesis that $\beta_o = o$ at $\alpha = .05$. Hence it appears for most strata that a separate ratio estimator as opposed to a regression estimator is an appropriate model for estimation of current level.

Since each of the estimators attain similar precision on estimation of the total no clearcut choice of the regression, separate ratio or combined ratio estimators can be made at this time. However, we recommend computing all three estimators and comparing them to the direct expansion estimator. This comparison should be made on both level and precision. These additional estimators may provide extremely valuable information in off quarter analysis (March, September) since only the multiple frame estimate is computed and charted for these quarters.

FUTURE STUDY

The proposed rotation plan suggested has several attractive features which may provide the necessary framework to retain a portion of the sample even after list update (usually in November). The "adds" and "deletes" from the list could be sampled as well as operations that change between strata. These components of a more complex composite successive sampling estimator could then be used to make estimates of quarter to quarter change after an update to the list. Of course this type of estimation would require a complicated historical data base system, as is currently being designed into the List Sampling Frame.

Another possibility, to lessen the potential impact on the U.S. indication of an entirely new sample across all states for the December survey is to rotate the time of update for individual states. Possibly one fourth of the states should update their list prior to each quarter.

Further research into composite estimators similar in form to ones used by the Census Bureau [9,15,16] in their monthly retail trade surveys is recommended. The research conducted in this study investigated three frequently used successive sampling estimators. No estimator clearly distinquished itself as "best" in this study. However, all estimators were more precise than the direct expansion estimator.

CONCLUSIONS/RECOMMENDATIONS

1) With the proposed rotation sampling plan, all strata (except the largest E.O.) are rotated uniformly between quarters. This will promote consistency in the estimates over time at both the state and national level. It will end the variety of rotation plans currently being used in the Agency (see Table A1). We recommend that 50 percent of the units be retained between quarters and that all states in the 14 hog multiple frame states be put on this rotation plan. The estimators also provide additional information in the critical March and September survey periods when only the multiple frame estimate is computed.

2) Quantitative response burden is limited to two contacts per year, instead of the up to four contacts in some strata, with guaranteed overlap for part of the sample in three of the four quarters. This takes advantage of the high correlation in survey data between quarters so that everyone from the previous quarter need not be recontacted.

3) Double sampling regression or ratio estimators provide composite estimates with similar or smaller variance than the direct expansion estimate. The sample design still allows for direct expansion estimates to be computed with the same summary procedures while also permitting the computation of composite estimates for comparison purposes. A weight of one-half should be used to combine the matched and unmatched portions of the composite estimates.

References

1. Auadhoni, M.S., Sukhatme, B.V., "Estimation in Sampling on Two Successive Occasions," Statistica Neerlandica, Vol 26, No 2, 1972.
2. Auadheni, M.S., Sukhatme, B.V., "A Comparison of Two Sampling Procedures with an Application to Successive Sampling," Journal Royal Stat. Soc., Series C, Vol 19, pp 251-259.
3. Bailar, Barbara, "The Effects of Rotation Group Bias on Estimates from Panel Surveys," JASA, Vol 70, no 349, pp 23-30.
4. Cochran, W., Sampling Techniques, John Wiley & Son, New York, N.Y., 1963.
5. DeGraft-Johnson, "Some Contributions to the Theory of Two Phase Sampling," Ph.D. Dissertation, Iowa State University, Ames, Iowa, 1969.
6. Eckler, A.R., "Rotation Sampling," Annuals of Math Statistics, Vol 26, pp 664-685.
7. Ford, Barry L., "Missing Data Procedures: A Comparative Study (Part 2)," Sample Survey Research Branch, ESCS, USDA, Washington, D.C., 20250, June 1978.
8. Gleason, C.P., and Tortora, R.D., "Successive Sampling of Two Overlapping Frames," Proceedings of the Joint Statistical Meetings, San Diego, Cal., August 1978.
9. Isaki, Cary T., Wolter, Kirk M., Sturdevant, Tyler R., Monsour, Nash J., and Trager, Mitchell L., "Sample Redesign of the Census Bureau's Monthly Business Surveys," Proceeding of the Business and Economics Statistics Section, ASA, 1976, Boston, Mass.
10. Krishman, P., "Rotation Sampling with a Fixed Panel Using Various Types of Difference Estimators," Symposium on Survey Sampling, Univ. of N.C., Chapel Hill, N.C., April 14-17, 1977.

11. Patterson, H.D., "Sampling on Successive Occasions with Partial Replacement of Units," Journal Royal Stat. Soc., B 12, pp 241-255, 1950.

12. Rao, J.N.K., Graham, J.E., "Rotation Designs for Sampling on Repeated Occasions," JASA, Vol 59, pp 492-509, 1964.

13. Sen, A.R., Sellers, S., Smith, G.E.J., "The Use of a Ratio Estimate in Successive Sampling," Biometrics, Vol 51, pp 673-684, 1975.

14. Singh, D., "Estimates in Successive Sampling Using a Multi-Stage Design," JASA, Vol 63, pp 99-112, 1968.

15. Wolter, Kirk M., Rsaki, Cary T., Sturdevant, Tyler R., Monsour, Nash J., and Mayes, Fred M., "Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys," Unpublished Census Bureau Report.

16. Woodruff, Ralph S., "The Use of Rotating Samples in the Census Bureau's Monthly Surveys," JASA, Vol 58, pp 454-467, 1963.

Appendix A

Table A1: List Stratum Rotation Plans by State - Hog M.F. 1978

State	Plan 1 One Sample, 4 Contacts Stratum No.	Plan 2 Two Samples, 2 Contacts Stratum No.	Plan 3 Two Samples, 2 Contacts Stratum No.	Plan 4 Four Samples, 1 Contact Stratum No.
Georgia	81 - 84	95		93 - 94
Illinois	98	81 - 86, 93		
Indiana	84 - 87	94		81 - 83, 93
Iowa	82 - 86, 94		81 & 93	
Kansas	81, 82	94		83 - 93
Kentucky	All Strata			
Minnesota	81 - 88	95		93 - 94
Missouri	81 - 83	84 - 86, 94		93
Nebraska	86, 88	81 - 85		95
N. Carolina	81 - 85	95		93 - 94
Ohio	81 - 88			93
S. Dakota		All Strata		
Texas		All Strata		
Wisconsin				All Strata

Table A2: Number of samples by survey period and number of merged records by survey period.

Iowa Dec/March 1978

Stratum	Description	Sample Size		% OL Between Survey Periods
		Dec	March	
1	No Hogs	395	396	0
2	1-199	272	272	100
3	200-399	220	222	99
4	400-599	220	220	100
5	600-999	269	273	99
6	1000-1999	167	172	99
805	2000-3999	171	171	0
806	4000-9999	78	78	100
905	10,000+	9	9	100

Iowa June/September 1976

Stratum	Description	Sample Size		% OL Between Survey Periods
		Dec	March	
1	No Livestock	303	304	0
2	No Hogs	223	222	0
3	1-199	239	239	100
4	200-349	225	225	100
5	350-599	283	283	100
6	600-999	225	225	100
7	1000-1999	138	138	100
85	2000-8499	113	113	0
95	85+	20	21	99

Minnesota June/Sept 1977

Stratum	Description	Sample Size June	Sample Size Sept	% OL Between Survey Periods
11	No Info	68	68	100
12	No Livestock	358	358	100
31	No Hogs	228	228	100
61	1-74	362	362	100
62	75-149	273	273	100
63	150-299	208	208	100
64	300-399	155	155	100
65	400-699	141	141	100
85	700-999	80	81	0
86	1000-2499	74	73	0
95	2500+	14	14	100

Kansas Dec/March 1977 1/

Stratum	Description	Sample Size Dec	Sample Size March	% OL Between Survey Periods
14	1-99	348	348	100
15	100-199	269	269	100
16	200-599	233	233	100
85	600-999	88	81	0
86	1000-999	64	64	0
95	4000+	24	24	100

1/ Kansas, stratum 10 and 12, subsample 0's from the December Survey and used all positives from December in the March Sample. They actually formulate two n strata (11 and 13) in March. For this analysis we used just the comparable strata.

Table A3: Matched sample sizes for all data sets in the analysis

Iowa Dec./March 1978

Stratum	No. Units Responding Both Times	Matched Sample Size
1	0	0
2	209	115
3	145	74
4	138	75
5	184	97
6	108	56
805	0	0
806	69	36
905	9	0*

Iowa June/Sept 1976

Stratum	No. Units Responding Both Times	Match Sample Size
1	0	0
2	0	0
3	188	94
4	162	84
5	170	87
6	137	74
7	86	51
85	0	0
95	21	0*

Minnesota June/Sept 1977

Stratum	No. Units Responding Both Times	Matched Sample Size
11	63	31
12	332	157
31	210	111
61	306	153
62	207	96
63	143	69
64	111	47
65	92	50
85	0	0
86	0	0*
95	14	0*

Kansas Dec/Mar 1977

Stratum	No. Units Responding Both Times	Match Sample Size
14	263	137
15	186	98
16	168	68
85	0	0
86	0	0
95	24	0*

* Strata with 100 percent coverage was not subsampled

Table A4 : Mean/Unit, Regression and Ratio Estimates

Iowa Dec/March 1978

Stratum	Mean/Unit	Regression	Ratio
2	47.92	49.42	52.65
3	130.99	132.00	134.41
4	214.65	213.72	213.64
5	359.07	358.55	358.62
6	534.34	528.99	529.22
806	1919.47	1821.05	1806.09

Iowa June/Sept 1976

Stratum	Mean/Unit	Regression	Ratio
3	47.74	47.84	48.00
4	136.06	136.07	135.62
5	239.35	244.58	246.26
6	343.26	356.18	360.20
7	546.17	537.83	538.34

Minnesota June/Sept 1977

Stratum	Mean/Unit	Regression	Ratio
11	4.03	2.99	2.85
12	5.18	5.46	5.28
31	2.22	2.19	2.12
61	26.56	27.36	27.76
62	62.15	61.38	61.52
63	161.20	161.63	161.84
64	226.02	232.60	233.98
65	388.69	399.48	402.02

Kansas Dec/March 1977

Stratum	Mean/Unit	Regression	Ratio
14	45.27	47.54	48.10
15	122.31	125.36	126.26
16	274.61	278.02	280.87

Table A5: Gain in precision (in %) of regression and ratio over the direct expansion estimate.

Iowa Dec/March 1978

Stratum	Gain Reg Over Mean	Gain Ratio Over Mean	Gain Reg Over Ratio
2	4.1	-5.4	10.1
3	3.6	-1.5	5.2
4	23.1	22.8	.2
5	28.3	26.5	1.4
6	15.3	14.5	.8
806	37.4	34.1	2.4

Iowa June/Sept 1971

Stratum	Gain Reg Over Mean	Gain Ratio Over Mean	Gain Reg Over Ratio
3	16.8	15.3	1.3
4	23.0	20.2	2.4
5	21.2	19.6	1.4
6	30.6	29.3	1.0
7	30.7	30.4	.2

Minnesota June/Sept 1977

Stratum	Gain Reg Over Mean	Gain Ratio Over Mean	Gain Reg Over Ratio
11	50.4	47.9	1.7
12	44.7	44.7	0.0
31	1.9	-11.9	15.7
61	26.9	22.9	3.2
62	23.2	20.7	2.1
63	33.3	32.5	.6
64	20.1	17.2	2.3
65	29.9	28.8	.8

Kansas Dec/March 1977

Stratum	Gain Reg Over Mean	Gain Ratio Over Mean	Gain Reg Over Ratio
14	20.3	19.7	.4
15	19.9	19.0	.7
16	19.9	17.3	2.2

Table A6: Statistics for Comparison of Weights to Combine Matched and Unmatched Estimates of Totals

Stratum	Matched Sample Size M	Correlation Coefficient ρ	Ratio of C.V.'s CV_1/CV_2 Δ	Ratio of Sample Sizes N_1/N_2 θ	Weights Matched Portion of Estimator: Separate Ratio Weight	Combined Ratio Weight	Separate Regression Weight
	IOWA DEC/MAR 78						
2	115	.38	1.01	1.02	.47	.63	.52
3	74	.36	.80	1.05	.43	.63	.46
4	75	.77	.83	1.06	.56	.63	.56
5	97	.81	.97	1.11	.59	.63	.60
6	56	.67	.81	1.04	.53	.63	.53
806	36	.93	1.14	.99	.62	.63	.63
	KANSAS DEC/MARCH 77						
14	137	.75	.85	1.00	.56	.49	.56
15	98	.74	.88	1.02	.56	.49	.56
16	68	.73	.95	1.05	.46	.49	.47
	MINNESOTA JUNE/SEPT 77						
11	31	.99	1.15	.98	.65	.56	.65
12	157	.96	.95	.99	.62	.56	.62
31	111	.28	1.13	.96	.43	.56	.51
61	153	.84	1.10	.97	.57	.56	.58
62	96	.79	1.01	.97	.52	.56	.53
63	69	.86	.96	1.06	.58	.56	.58
64	47	.74	.97	1.02	.48	.56	.49
65	50	.83	.95	1.11	.62	.56	.62

Table A7: Estimates of Regression Coefficients and t-Test of Hypothesis H_o: $\beta_o = 0$

Stratum	Estimate of Parameters β_o	β_1	t-Test H_o: $\beta_o = 0$, $\alpha = .05$
	Iowa Dec/Mar 1978		
2	12.27	.65	NS
3	61.70	.37	Significant
4	39.40	.76	NS
5	55.88	.87	Significant
6	122.03	.70	NS
806	391.72	.82	Significant
	Kansas Dec/Mar 1977		
14	5.70	.79	NS
15	17.32	.76	NS
16	63.70	.73	Significant
	Minnesota June/Sept 1977		
11	.27	.63	NS
12	-.07	.81	NS
31	1.66	1.05	NS
61	8.38	.83	Significant
62	11.73	.75	NS
63	21.85	.87	NS
64	56.00	.99	NS
65	69.67	.87	NS

NS = Not Significant

Appendix B: Estimation Formulas for Regression, Ratio and Combined Ratio Estimate

To develop this theory we summarize the notation as developed in Sen, et al. [13]. Within a list frame stratum let,

n_1 = total sample size on first occasion

n_2 = total sample size on second occasion

m = matched sample size of those operations who reported on both occasions.

u_1 = sample size of those who responded only on the first time

u_2 = sample size of those who responded only on the second occasion, So, $u_1 = n_1 - m$ and $u_2 = n_2 - m$;

$\bar{y}_1$ = total sample mean on first occasion

$\bar{y}_2$ = total sample mean on second occasion

$\bar{y}_{1m}$ = matched sample mean on first occasion

$\bar{y}_{2m}$ = matched sample mean on second occasion

$\bar{y}_{1u}$ = unmatched sample mean on first occasion

$\bar{y}_{2u}$ = unmatched sample mean on second occasion

S_1^2 = population variance on first occasion

S_2^2 = population variance on second occasion

Δ = $(S_1/\bar{Y}_1)/(S_2/\bar{Y}_2)$; ratio of the C.V. on the first occasion to that on the second.

ρ = correlation coefficient between y_1 and y_2.

$Z = \Delta\ (2\rho - \Delta)$

$\lambda = m/n_2$

$\Theta = n_1/n_2$

So for example, n_1 is the total number of useable reports for a particular state and stratum. For purposes of this analysis we will ignore finite population corrections in comparing the regression, ratio and mean/unit estimates (direct expansion).

Composite Ratio Estimation Within Strata

The matched (m units) and unmatched units on second occasion provide independent estimates ($\bar{y}_{2m}$ and $\bar{y}_{2u}$) of the population mean $\bar{Y}_2$ on the second occasion. For the matched portion of the sample and improved estimate $\bar{y}'_{2m}$ of $\bar{Y}_2$ is the double sampling ratio estimate

$$\bar{y}'_{2m} = \frac{\bar{y}_{2m}}{\bar{y}_{1m}} \quad \bar{y}_1 \quad (1)$$

The variance of $\bar{y}'_{2m}$ is,

$$V(\bar{y}'_{2m}) = S_2^2 \frac{(n_1 - u_1 Z)}{m\, n_1} \quad (2).$$

As an estimate of the variance we can replace $\hat{S}_2$ and Z by their appropriate sample estimates. Assuming $\Delta = 1$ (C.V.'s between survey periods are equal. This is not an unreasonable assumption for the quarterly hog survey.), it is necessary for $\rho > .5$ for $\bar{y}'_{2m}$ to be a gain over $\bar{y}_2$ (Cochran p. 165). A ratio estimate, $\bar{y}_{2r}$ of the mean $\bar{Y}_2$ of the population on the second occasion is given by combining the independent estimates $\bar{y}'_{2m}$ and $\bar{y}_{2u}$ as

$$\bar{y}_{2r} = w\, \bar{y}'_{2m} + (1 - w)\, \bar{y}_{2u} \quad (3).$$

The best estimate of the mean $\bar{Y}_2$ is obtained by using values of w which minimize $V(\bar{y}_{2r})$. Taking the derivative of (3) with respect to w gives the value which minimizes $V(\bar{y}_{2r})$, which is

$$w = V(\bar{y}_{2u})/(V(\bar{Y}_{2u}) + V(\bar{y}'_{2m})). \quad (4)$$

With this value for w the $V(\bar{y}_{2r})$ is

$$V(\bar{y}_{2r}) = \frac{S^2_{y_2} \{n_1 - u_1 \, z\}}{n_1 \, n_2 - z \, u_1 \, u_2} \, . \qquad (5)$$

The gain of the ratio estimate over the mean/unit estimate (commonly called a direct expansion estimator) is

$$G_1 = \{V(\bar{y}_2) - V(\bar{y}_{2r})\}/V(\bar{y}_{2r}).$$

$$= \frac{z \, \lambda \, (\Theta - \lambda)}{\Theta - z \, (\Theta - \lambda)} \qquad (6)$$

The optimum percentage to match for the ratio estimate can be found by taking the derivative of G_1 with respect to λ. This gives

$$\lambda_{\text{opt, ratio}} = \frac{\Theta\sqrt{1 - z}}{1 + \sqrt{1 - z}} \qquad (7)$$

To arrive at estimates of totals for the list frame each mean by strata was expanded by the population size N_h and summed over all strata. Similar computations were performed for variance estimation of the total.

Composite Combined Ratio Estimation

Instead of computing the estimator (1) on a stratum by stratum basis an alternative procedure is to compute a combined ratio over all strata where there were matched samples. To estimate the total on the second occasion this estimator becomes:

$$\hat{Y}_{RC} = w \left(\frac{\Sigma N_h \, \bar{y}_{2mh}}{\Sigma N_h \, \bar{y}_{1mh}}\right) \hat{Y}_1 + (1 - w) \; \Sigma N_h \; \bar{y}_{2uh}$$

where the last subscript represents the h-th stratum, and

N_h = population size in h-th stratum,

w is a weight which combines the matched and unmatched portion of the estimator,

$\hat{Y}_1$ is the previous survey's direct expansion estimate of the total for all the matched strata.

The variance of $\hat{Y}_{RC}$ is

$$V(\hat{Y}_{RC}) = [\Sigma N_h^2 \left(\frac{S^2_{y_{2h}} + R^2 S^2_{y_{1h}} - 2 R \rho_h S_{y_{1h}} S_{y_{2h}}}{m_a}\right) +$$

$$(2 R S_{y_{1h}} S_{y_{2h}} - R S^2_{y_{1h}})] w^2 + (1 - w)^2 \Sigma N_h^2 S^2_{y_{2h}} / u_2$$

where $R = \frac{\Sigma N_h \bar{y}_{2mh}}{\Sigma N_h \bar{y}_{1mh}}$ and

$S^2_{y_{1h}}$, $S^2_{y_{2h}}$ are the within stratum population variances on the 1^{st} and 2^{nd} occasion, respectively.

Composite Regression Estimation

Another estimate of the mean/unit in each stratum is the regression estimate. Patterson [11] was the first to suggest such an estimate. The regression estimate can be written (as was the ratio) as a combination of the matched with the unmatched with an optimum weight t to minimize $V(\bar{y}_{2,reg})$ as follows:

$$\bar{y}_{2\,reg} = \frac{t\,\{\bar{y}_{2m} + \rho\, S_{y_2}/S_{y_1}\, (\bar{y}_1 - \bar{y}_{1m})\} + u_2\, \bar{y}_{2u}}{t + u_2}$$

where $t = 1/(\frac{\rho^2}{n_1} + \frac{(1 - \rho^2)}{m})$ (8)

and u_2 is the unmatched sample size.

The variance of $\bar{y}_{2,reg}$ is

$$V(\bar{y}_{2,reg}) = \frac{S^2_{y_2}\, (n_1 - \rho^2\, u_1)}{n_1\, n_2 - \rho^2\, u_1\, u_2} \qquad (9)$$

The gain of the regression estimate over the ratio estimate is,

$$G_2 = (V(\bar{y}_{2r}) - V(\bar{y}_{2,reg}))/V(\bar{y}_{2,reg}) \quad (10)$$

and the gain of the regression over the mean per unit estimate is

$$G_3 = (V(\bar{y}_2) - V(\bar{y}_{2,reg}))/V(\bar{y}_{2,reg}) \quad (11).$$

G_2 is usually small except where Δ (the ratio of the C.V.'s) is large. For our particular application the gains (in variance) were not large for any state/stratum combination. The optimum percentage to match for the regression estimate is

$$\lambda_{opt,\ reg} = \frac{\theta\sqrt{1 - \rho^2}}{1 + \sqrt{1 - \rho^2}} \qquad (12).$$

This generalizes Cochran's results [4] for unequal sample sizes on each occasion. Generally the optimum percentage to match for the regression and the ratio estimate are not the same. However, for the quarterly hog survey the ratio of the C.V.'s between survey periods is nearly 1, (see Table A6 in Appendix A). Therefore, $\Delta = 1$, implies $\theta = 2\rho - 1$ and

$\lambda_{opt,\ ratio}$ becomes

$$\lambda_{opt,\ ratio} = \frac{\Theta\sqrt{2(1 - \rho)}}{1 + \sqrt{2(1 - \rho)}}$$

It can be shown that $\lambda_{opt,\ ratio} > \lambda_{opt,\ reg}$ when $\Delta = 1$, therefore, the ratio estimate always requires a larger matching percentage.